OIPE

RAW SEQUENCE LISTING   DATE: 08/12/2004
PATENT APPLICATION: **US/09/774,954**   TIME: 18:36:35

Input Set : **N:\Crf3\RULE60\09774954.raw**
Output Set: **N:\CRF4\08122004\I774954.raw**

SEQUENCE LISTING

```
 3 (1) GENERAL INFORMATION:
 5       (i) APPLICANT: Yang Wang, Michael W. Spellman
 7      (ii) TITLE OF INVENTION: O-Fucosyltransferase
 9     (iii) NUMBER OF SEQUENCES: 17
11      (iv) CORRESPONDENCE ADDRESS:
12           (A) ADDRESSEE: Genentech, Inc.
13           (B) STREET: 1 DNA Way
14           (C) CITY: South San Francisco
15           (D) STATE: California
16           (E) COUNTRY: USA
17           (F) ZIP: 94080
19       (v) COMPUTER READABLE FORM:
20           (A) MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
21           (B) COMPUTER: IBM PC compatible
22           (C) OPERATING SYSTEM: PC-DOS/MS-DOS
23           (D) SOFTWARE: WinPatin (Genentech)
25      (vi) CURRENT APPLICATION DATA:
C--> 26          (A) APPLICATION NUMBER: US/09/774,954
C--> 27          (B) FILING DATE: 30-Jan-2001
28           (C) CLASSIFICATION:
30     (vii) PRIOR APPLICATION DATA:
W--> 31          (A) APPLICATION NUMBER: US/08/978,741
32           (B) FILING DATE: 26-NOV-1997
W--> 34          (A) APPLICATION NUMBER: 08/792,498
35           (B) FILING DATE: 31-JAN-1997
37    (viii) ATTORNEY/AGENT INFORMATION:
38           (A) NAME: Svoboda, Craig G.
39           (B) REGISTRATION NUMBER: 39,044
40           (C) REFERENCE/DOCKET NUMBER: P1041P1
42      (ix) TELECOMMUNICATION INFORMATION:
43           (A) TELEPHONE: 650/225-1489
44           (B) TELEFAX: 650/952-9881
45 (2) INFORMATION FOR SEQ ID NO: 1:
47       (i) SEQUENCE CHARACTERISTICS:
48           (A) LENGTH: 1514 base pairs
49           (B) TYPE: Nucleic Acid
50           (C) STRANDEDNESS: Single
51           (D) TOPOLOGY: Linear
53      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:
56 ATGCCCGCGG GCTCCTGGGA CCCGGCCGGT TACCTGCTCT ACTGCCCCTG 50
58 CATGGGGCGC TTTGGGAACC AGGCCGATCA CTTCTTGGGC TCTCTGGCAT 100
60 TTGCAAAGCT GCTAAACCGT ACCTTGGCTG TCCCTCCTTG GATTGAGTAC 150
```

RAW SEQUENCE LISTING   DATE: 08/12/2004
PATENT APPLICATION: US/09/774,954   TIME: 18:36:35

Input Set : N:\Crf3\RULE60\09774954.raw
Output Set: N:\CRF4\08122004\I774954.raw

```
 62  CAGCATCACA AGCCTCCTTT CACCAACCTC CATGTGTCCT ACCAGAAGTA  200
 64  CTTCAAGCTG GAGCCCCTCC AGGCTTACCA TCGGGTCATC AGCTTGGAGG  250
 66  ATTTCATGGA GAAGCTGGCA CCCACCCACT GGCCCCCTGA GAAGCGGGTG  300
 68  GCATACTGCT TTGAGGTGGC AGCCCAGCGA AGCCCAGATA AGAAGACGTG  350
 70  CCCCATGAAG GAAGGAAACC CCTTTGGCCC ATTCTGGGAT CAGTTTCATG  400
 72  TGAGTTTCAA CAAGTCGGAG CTTTTTACAG GCATTTCCTT CAGTGCTTCC  450
 74  TACAGAGAAC AATGGAGCCA GAGATTTTCT CCAAAGGAAC ATCCGGTGCT  500
 76  TGCCCTGCCA GGAGCCCCAG CCCAGTTCCC CGTCCTAGAA GAACACAGGC  550
 78  CACTACAGAA GTACATGGTA TGGTCAGACG AAATGGTGAA GACGGGAGAG  600
 80  GCCCAGATTC ATGCCCACCT TGTCCGGCCC TATGTGGGCA TTCATCTGCG  650
 82  CATTGGCTCT GACTGGAAGA ACGCCTGTGC CATGCTGAAG GACGGGACTG  700
 84  CAGGCTCGCA CTTCATGGCC TCTCCGCAGT GTGTGGGCTA CAGCCGCAGC  750
 86  ACAGCGGCCC CCCTCACGAT GACTATGTGC CTGCCTGACC TGAAGGAGAT  800
 88  CCAGAGGGCT GTGAAGCTCT GGGTGAGGTC GCTGGATGCC CAGTCGGTCT  850
 90  ACGTTGCTAC TGATTCCGAG AGTTATGTGC CTGAGCTCCA ACAGCTCTTC  900
 92  AAAGGGAAGG TGAAGGTGGT GAGCCTGAAG CCTGAGGTGG CCCAGGTCGA  950
 94  CCTGTACATC CTCGGCCAAG CCGACCACTT TATTGGCAAC TGTGTCTCCT  1000
 96  CCTTCACTGC CTTTGTGAAG CGGGAGCGGG ACCTCCAGGG GAGGCCGTCT  1050
 98  TCTTTCTTCG GCATGGACAG GCCCCCTAAG CTGCGGGACG AGTTCTGATT  1100
100  CTGGCCGGAG CACCAGACCC TCTGATCCTG GAGGGACCAG AGTCTGAGCT  1150
102  GGTCCTTCCA GCCAGGCCTG GCAGCCAGAG GTGCTCCGGG ATTGCAAACT  1200
104  CCTCTTCTCA CCTGCCAAAG ATGGAGAAGA GTGCCAGGGA CCCCTCAAGG  1250
106  AGGGAGACGC TCCATATCCC AGGGCATAGG ACTTGCAGGT TCCTAGGAGC  1300
108  AGGAGCATCT CCCATCGCAC GTGCTTTCTG CTCTTCTGGG AATTTCTCAC  1350
110  ACTGGCAAAG CAGTCCAGCC TCCGTCTTCT GGTCCACTCT GCTCTGAGCA  1400
112  GCCTGGGATG CTGAACTCTT CAGAGAGATT TTTTTATAGA GAGATTTCTA  1450
114  TAATTTTGAT ACAAGGTCAT GACTATCCTA GAACTCTCTG TGGTTTTTGA  1500
116  AAATCATTGA ATTC  1514
118  (2) INFORMATION FOR SEQ ID NO: 2:
120      (i) SEQUENCE CHARACTERISTICS:
121          (A) LENGTH: 365 amino acids
122          (B) TYPE: Amino Acid
123          (D) TOPOLOGY: Linear
125      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:
127  Met Pro Ala Gly Ser Trp Asp Pro Ala Gly Tyr Leu Leu Tyr Cys
128   1               5                  10                  15
130  Pro Cys Met Gly Arg Phe Gly Asn Gln Ala Asp His Phe Leu Gly
131               20                  25                  30
133  Ser Leu Ala Phe Ala Lys Leu Leu Asn Arg Thr Leu Ala Val Pro
134               35                  40                  45
136  Pro Trp Ile Glu Tyr Gln His His Lys Pro Pro Phe Thr Asn Leu
137               50                  55                  60
139  His Val Ser Tyr Gln Lys Tyr Phe Lys Leu Glu Pro Leu Gln Ala
140               65                  70                  75
142  Tyr His Arg Val Ile Ser Leu Glu Asp Phe Met Glu Lys Leu Ala
143               80                  85                  90
145  Pro Thr His Trp Pro Pro Glu Lys Arg Val Ala Tyr Cys Phe Glu
146               95                 100                 105
148  Val Ala Ala Gln Arg Ser Pro Asp Lys Lys Thr Cys Pro Met Lys
```

RAW SEQUENCE LISTING   DATE: 08/12/2004
PATENT APPLICATION: US/09/774,954   TIME: 18:36:35

Input Set : N:\Crf3\RULE60\09774954.raw
Output Set: N:\CRF4\08122004\I774954.raw

```
149                   110               115              120
151  Glu Gly Asn Pro Phe Gly Pro Phe Trp Asp Gln Phe His Val Ser
152                   125               130              135
154  Phe Asn Lys Ser Glu Leu Phe Thr Gly Ile Ser Phe Ser Ala Ser
155                   140               145              150
157  Tyr Arg Glu Gln Trp Ser Gln Arg Phe Ser Pro Lys Glu His Pro
158                   155               160              165
160  Val Leu Ala Leu Pro Gly Ala Pro Ala Gln Phe Pro Val Leu Glu
161                   170               175              180
163  Glu His Arg Pro Leu Gln Lys Tyr Met Val Trp Ser Asp Glu Met
164                   185               190              195
166  Val Lys Thr Gly Glu Ala Gln Ile His Ala His Leu Val Arg Pro
167                   200               205              210
169  Tyr Val Gly Ile His Leu Arg Ile Gly Ser Asp Trp Lys Asn Ala
170                   215               220              225
172  Cys Ala Met Leu Lys Asp Gly Thr Ala Gly Ser His Phe Met Ala
173                   230               235              240
175  Ser Pro Gln Cys Val Gly Tyr Ser Arg Ser Thr Ala Ala Pro Leu
176                   245               250              255
178  Thr Met Thr Met Cys Leu Pro Asp Leu Lys Glu Ile Gln Arg Ala
179                   260               265              270
181  Val Lys Leu Trp Val Arg Ser Leu Asp Ala Gln Ser Val Tyr Val
182                   275               280              285
184  Ala Thr Asp Ser Glu Ser Tyr Val Pro Glu Leu Gln Gln Leu Phe
185                   290               295              300
187  Lys Gly Lys Val Lys Val Val Ser Leu Lys Pro Glu Val Ala Gln
188                   305               310              315
190  Val Asp Leu Tyr Ile Leu Gly Gln Ala Asp His Phe Ile Gly Asn
191                   320               325              330
193  Cys Val Ser Ser Phe Thr Ala Phe Val Lys Arg Glu Arg Asp Leu
194                   335               340              345
196  Gln Gly Arg Pro Ser Ser Phe Phe Gly Met Asp Arg Pro Pro Lys
197                   350               355              360
199  Leu Arg Asp Glu Phe
200                   365
202  (2) INFORMATION FOR SEQ ID NO: 3:
204        (i) SEQUENCE CHARACTERISTICS:
205            (A) LENGTH: 61 amino acids
206            (B) TYPE: Amino Acid
207            (D) TOPOLOGY: Linear
209        (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:
W--> 211  Arg Leu Ala Gly Ser Trp Asp Leu Ala Gly Tyr Leu Leu Tyr Xaa
212    1               5               10               15
214  Pro Xaa Met Gly Arg Phe Gly Asn Gln Ala Asp His Phe Leu Gly
215                   20               25               30
217  Ser Leu Ala Phe Ala Lys Leu Xaa Val Arg Thr Leu Ala Val Pro
218                   35               40               45
220  Pro Trp Ile Glu Tyr Gln His His Lys Pro Pro Phe Thr Asn Leu
221                   50               55               60
```

RAW SEQUENCE LISTING   DATE: 08/12/2004
PATENT APPLICATION: US/09/774,954   TIME: 18:36:35

Input Set : N:\Crf3\RULE60\09774954.raw
Output Set: N:\CRF4\08122004\I774954.raw

```
223  His
224  61
226  (2) INFORMATION FOR SEQ ID NO: 4:
228      (i) SEQUENCE CHARACTERISTICS:
229          (A) LENGTH: 1300 base pairs
230          (B) TYPE: Nucleic Acid
231          (C) STRANDEDNESS: Single
232          (D) TOPOLOGY: Linear
234      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:
237  TTATTCATAC CGTCCCACCA TCGGGCGCGG ATCAGATCCA TGGCCAAGTT 50
239  CCTGGTCAAC GTGGCCCTGC TGCTGCTGCT GCTGCTGCTG TCCGGAGCCT 100
241  GGGCCCATAT GAGATCCCAT CACCATCACC ATCACATGCC CGCGGGCTCC 150
243  TGGGACCCGG CCGGTTACCT GCTCTACTGC CCCTGCATGG GGCGCTTTGG 200
245  GAACCAGGCC GATCACTTCT TGGGCTCTCT GGCATTTGCA AAGCTGCTAA 250
247  ACCGTACCTT GGCTGTCCCT CCTTGGATTG AGTACCAGCA TCACAAGCCT 300
249  CCTTTCACCA ACCTCCATGT GTCCTACCAG AAGTACTTCA AGCTGGAGCC 350
251  CCTCCAGGCT TACCATCGGG TCATCAGCTT GGAGGATTTC ATGGAGAAGC 400
253  TGGCACCCAC CCACTGGCCC CCTGAGAAGC GGGTGGCATA CTGCTTTGAG 450
255  GTGGCAGCCC AGCGAAGCCC AGATAAGAAG ACGTGCCCCA TGAAGGAAGG 500
257  AAACCCCTTT GGCCCATTCT GGGATCAGTT TCATGTGAGT TTCAACAAGT 550
259  CGGAGCTTTT TACAGGCATT TCCTTCAGTG CTTCCTACAG AGAACAATGG 600
261  AGCCAGAGAT TTTCTCCAAA GGAACATCCG GTGCTTGCCC TGCCAGGAGC 650
263  CCCAGCCCAG TTCCCCGTCC TAGAGGAACA CAGGCCACTA CAGAAGTACA 700
265  TGGTATGGTC AGACGAAATG GTGAAGACGG GAGAGGCCCA GATTCATGCC 750
267  CACCTTGTCC GGCCCTATGT GGGCATTCAT CTGCGCATTG GCTCTGACTG 800
269  GAAGAACGCC TGTGCCATGC TGAAGGACGG GACTGCAGGC TCGCACTTCA 850
271  TGGCCTCTCC GCAGTGTGTG GGCTACAGCC GCAGCACAGC GGCCCCCCTC 900
273  ACGATGACTA TGTGCCTGCC TGACCTGAAG GAGATCCAGA GGGCTGTGAA 950
275  GCTCTGGGTG AGGTCGCTGG ATGCCCAGTC GGTCTACGTT GCTACTGATT 1000
277  CCGAGAGTTA TGTGCCTGAG CTCCAACAGC TCTTCAAAGG GAAGGTGAAG 1050
279  GTGGTGAGCC TGAAGCCTGA GGTGGCCCAG GTCGACCTGT ACATCCTCGG 1100
281  CCAAGCCGAC CACTTTATTG GCAACTGTGT CTCCTCCTTC ACTGCCTTTG 1150
283  TGAAGCGGGA GCGGGACCTC CAGGGGAGGC CGTCTTCTTT CTTCGGCATG 1200
285  GACAGGCCCC CTAAGCTGCG GGACGAGTTC TGATTCTGGC CGGAGCACCA 1250
287  GACCCTCTGA TCCTGGAGGG ACCAGAGTCT GAGCTGGTCC TTCCAGCCAG 1300
289  (2) INFORMATION FOR SEQ ID NO: 5:
291      (i) SEQUENCE CHARACTERISTICS:
292          (A) LENGTH: 11284 base pairs
293          (B) TYPE: Nucleic Acid
294          (C) STRANDEDNESS: Single
295          (D) TOPOLOGY: Linear
297      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:
300  AAGCTTTACT CGTAAAGCGA GTTGAAGGAT CATATTTAGT TGCGTTTATG 50
302  AGATAAGATT GAAAGCACGT GTAAAATGTT TCCCGCGCGT TGGCACAACT 100
304  ATTTACAATG CGGCCAAGTT ATAAAGATT CTAATCTGAT ATGTTTTAAA 150
306  ACACCTTTGC GGCCCGAGTT GTTTGCGTAC GTGACTAGCG AAGAAGATGT 200
308  GTGGACCGCA GAACAGATAG TAAAACAAAA CCCTAGTATT GGAGCAATAA 250
310  TCGATTTAAC CAACACGTCT AAATATTATG ATGGTGTGCA TTTTTTGCGG 300
312  GCGGGCCTGT TATACAAAAA AATTCAAGTA CCTGGCCAGA CTTTGCCGCC 350
```

RAW SEQUENCE LISTING   DATE: 08/12/2004
PATENT APPLICATION: US/09/774,954   TIME: 18:36:35

Input Set : **N:\Crf3\RULE60\09774954.raw**
Output Set: **N:\CRF4\08122004\I774954.raw**

```
314   TGAAAGCATA GTTCAAGAAT TTATTGACAC GGTAAAAGAA TTTACAGAAA 400
316   AGTGTCCCGG CATGTTGGTG GGCGTGCACT GCACACACGG TATTAATCGC 450
318   ACCGGTTACA TGGTGTGCAG ATATTTAATG CACACCCTGG GTATTGCGCC 500
320   GCAGGAAGCC ATAGATAGAT TCGAAAAAGC CAGAGGTCAC AAAATTGAAA 550
322   GACAAAATTA CGTTCAAGAT TTATTAATTT AATTAATATT ATTTGCATTC 600
324   TTTAACAAAT ACTTTATCCT ATTTTCAAAT TGTTGCGCTT CTTCCAGCGA 650
326   ACCAAAACTA TGCTTCGCTT GCTCCGTTTA GCTTGTAGCC GATCAGTGGC 700
328   GTTGTTCCAA TCGACGGTAG GATTAGGCCG GATATTCTCC ACCACAATGT 750
330   TGGCAACGTT GATGTTACGT TTATGCTTTT GGTTTTCCAC GTACGTCTTT 800
332   TGGCCGGTAA TAGCCGTAAA CGTAGTGCCG TCGCGCGTCA CGCACAACAC 850
334   CGGATGTTTG CGCTTGTCCG CGGGGTATTG AACCGCGCGA TCCGACAAAT 900
336   CCACCACTTT GGCAACTAAA TCGGTGACCT GCGCGTCTTT TTTCTGCATT 950
338   ATTTCGTCTT TCTTTTGCAT GGTTTCCTGG AAGCCGGTGT ACATGCGGTT 1000
340   TAGATCAGTC ATGACGCGCG TGACCTGCAA ATCTTTGGCC TCGATCTGCT 1050
342   TGTCCTTGAT GGCAACGATG CGTTCAATAA ACTCTTGTTT TTTAACAAGT 1100
344   TCCTCGGTTT TTTGCGCCAC CACCGCTTGC AGCGCGTTTG TGTGCTCGGT 1150
346   GAATGTCGCA ATCAGCTTAG TCACCAACTG TTTGCTCTCC TCCTCCCGTT 1200
348   GTTTGATCGC GGGATCGTAC TTGCCGGTGC AGAGCACTTG AGGAATTACT 1250
350   TCTTCTAAAA GCCATTCTTG TAATTCTATG GCGTAAGGCA ATTTGGACTT 1300
352   CATAATCAGC TGAATCACGC CGGATTTAGT AATGAGCACT GTATGCGGCT 1350
354   GCAAATACAG CGGGTCGCCC CTTTTCACGA CGCTGTTAGA GGTAGGGCCC 1400
356   CCATTTTGGA TGGTCTGCTC AAATAACGAT TTGTATTTAT TGTCTACATG 1450
358   AACACGTATA GCTTTATCAC AAACTGTATA TTTTAAACTG TTAGCGACGT 1500
360   CCTTGGCCAC GAACCGGACC TGTTGGTCGC GCTCTAGCAC GTACCGCAGG 1550
362   TTGAACGTAT CTTCTCCAAA TTTAAATTCT CCAATTTTAA CGCGAGCCAT 1600
364   TTTGATACAC GTGTGTCGAT TTTGCAACAA CTATTGTTTT TTAACGCAAA 1650
366   CTAAACTTAT TGTGGTAAGC AATAATTAAA TATGGGGGAA CATGCGCCGC 1700
368   TACAACACTC GTCGTTATGA ACGCAGACGG CGCCGGTCTC GGCGCAAGCG 1750
370   GCTAAAACGT GTTGCGCGTT CAACGCGGCA AACATCGCAA AAGCCAATAG 1800
372   TACAGTTTTG ATTTGCATAT TAACGGCGAT TTTTTAAATT ATCTTATTTA 1850
374   ATAAATAGTT ATGACGCCTA CAACTCCCCG CCCGCGTTGA CTCGCTGCAC 1900
376   CTCGAGCAGT TCGTTGACGC CTTCCTCCGT GTGGCCGAAC ACGTCGAGCG 1950
378   GGTGGTCGAT GACCAGCGGC GTGCCGCACG CGACGCACAA GTATCTGTAC 2000
380   ACCGAATGAT CGTCGGGCGA AGGCACGTCG GCCTCCAAGT GGCAATATTG 2050
382   GCAAATTCGA AAATATATAC AGTTGGGTTG TTTGCGCATA TCTATCGTGG 2100
384   CGTTGGGCAT GTACGTCCGA ACGTTGATTT GCATGCAAGC CGAAATTAAA 2150
386   TCATTGCGAT TAGTGCGATT AAAACGTTGT ACATCCTCGC TTTTAATCAT 2200
388   GCCGTCGATT AAATCGCGCA ATCGAGTCAA GTGATCAAAG TGTGGAATAA 2250
390   TGTTTTCTTT GTATTCCCGA GTCAAGCGCA GCGCGTATTT TAACAAACTA 2300
392   GCCATCTTGT AAGTTAGTTT CATTTAATGC AACTTTATCC AATAATATAT 2350
394   TATGTATCGC ACGTCAAGAA TTAACAATGC GCCCGTTGTC GCATCTCAAC 2400
396   ACGACTATGA TAGAGATCAA ATAAAGCGCG AATTAAATAG CTTGCGACGC 2450
398   AACGTGCACG ATCTGTGCAC GCGTTCCGGC ACGAGCTTTG ATTGTAATAA 2500
400   GTTTTTACGA AGCGATGACA TGACCCCCGT AGTGACAACG ATCACGCCCA 2550
402   AAAGAACTGC CGACTACAAA ATTACCGAGT ATGTCGGTGA CGTTAAAACT 2600
404   ATTAAGCCAT CCAATCGACC GTTAGTCGAA TCAGGACCGC TGGTGCGAGA 2650
406   AGCCGCGAAG TATGGCGAAT GCATCGTATA ACGTGTGGAG TCCGCTCATT 2700
408   AGAGCGTCAT GTTTAGACAA GAAAGCTACA TATTTAATTG ATCCCGATGA 2750
410   TTTTATTGAT AAATTGACCC TAACTCCATA CACGGTATTC TACAATGGCG 2800
```